

RESEARCH ARTICLE

Comparative transcriptomic analysis of genes in the triterpene saponin biosynthesis pathway in leaves and roots of *Ardisia kteniophylla* A. DC., a plant used in traditional Chinese medicine

Yuyang Lei^{1,2}  | AJ Harris¹  | Aihua Wang^{1,3}  | Liyun Zhao⁴  | Ming Luo^{5,6}  | Ji Li⁶ | Hongfeng Chen¹ 

¹Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

²Wuhan Guishan Mountain Scenic Management Office, Wuhan, China

³Key Laboratory of Environment Change and Resources Use in Beibu Gulf (Nanning Normal University), Ministry of Education, Nanning, China

⁴CAS Engineering Laboratory for Vegetation Ecosystem Restoration on Islands and Coastal Zones, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

⁵Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

⁶Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

Correspondence

Hongfeng Chen, Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China.

Email: h.f.chen@scbg.ac.cn

Funding information

This work was supported by the grants from Key Research and Development Program of Guangdong Province (2020B1111530004), the Forestry Flora and Fauna Conservation Project of Guangdong Province, China (E036011002)

Abstract

Ardisia kteniophylla (Primulaceae) is highly valued in traditional medicine due to its production of the pharmacologically active secondary metabolites, especially triterpenoid saponins in its roots. Although *A. kteniophylla* is very important in traditional medicine, the genetic basis for its production of triterpenoid saponins remains largely unknown. Therefore, we sequenced transcriptomes of *A. kteniophylla* to identify putative genes involved in production of triterpenoid saponins in both leaves and roots, and we used the transcriptomes to compare expression levels of these genes between the two organ systems. The production of triterpenoid saponins in plants is usually induced through hormonal signaling on account of the presence of pests. Thus, we treated plants with the hormones salicylic acid (SA) and methyl jasmonate (MeJA) and used quantitative real-time PCR (qRT-PCR) to investigate expression levels of genes involved in triterpenoid saponin biosynthesis. In total, we obtained transcriptomes for leaf and root tissues representing 52,454 unigenes. Compared with the leaf transcriptome, we found that 6092 unigenes were upregulated in the root, especially enzymes involved in the direct synthesis of triterpenoid saponins, while 6001 genes appeared downregulated, including those involved in precursory steps in the triterpenoid saponin biosynthesis pathway. Our results from qRT-PCR indicate that genes within the upstream parts of the triterpenoid saponin biosynthesis pathway may be upregulated under exposure to the applied hormones, but downstream genes are downregulated. This suggests possible conflicting effects of SA and MeJA in promoting the production of secondary metabolites on the one hand, and, on the other, limiting plant growth processes to devote energy to combating pests. We also performed an analysis of transcription factors (TFs) and found 997 unique transcripts belonging to 16 TF families. Our data may help to facilitate future work on triterpene saponins biosynthesis in *A. kteniophylla* with potential pharmacological and molecular breeding applications.

Yuyang Lei and AJ Harris contributed equally to this work and share first authorship.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

and Y936111001) and the Science and Technology Program of Guangdong Province, China (2019B1212006) made to Hongfeng Chen

KEYWORDS

DEGs, Illumina sequencing, Primulaceae, transcription factors, triterpenoid saponins

TAXONOMY CLASSIFICATION

Botany; Genomics

1 | INTRODUCTION

Ardisia kteniophylla A. DC. is a perennial plant within the Primulaceae family (Ericales; Huang et al., 2017), and its roots are widely known to accumulate pharmacologically active products such as saponins, benzoquinones, flavanones, and sterols (Sun, Li, et al., 2017; Sun, Jiang, et al., 2017; Zhou, 2017). These products likely form the biochemical basis for the utility of *A. kteniophylla* in traditional Chinese medicine. In particular, the dried root of *A. kteniophylla* is used in southern China for the treatment of rheumatism, muscle and bone pain, and traumatic injury (Dai et al., 2017; Tang, 2007; Xiang & Feng, 2002). The medicinal value of *A. kteniophylla* has driven annually increasing demand for the species and has, thus, placed pressure on wild resources, which can no longer adequately supply the markets. Wild populations are also threatened by large-scale illegal mining, which has resulted in considerable habitat loss. Therefore, unfortunately, *A. kteniophylla* is on the verge of extinction within China and globally due to both overharvesting and habitat destruction (Wei et al., 2015). While its conservation may be achievable based on comprehensive assessments of its genetic resources, so far, genetic data for *A. kteniophylla* are lacking.

The most pharmacologically significant bioactive compounds produced by *A. kteniophylla* are triterpenoid saponins (Mu et al., 2010), which consist of triterpene sapogenin, sugars, uronic acid, and other organic acids. Broadly, terpenes, such as triterpenoid saponins, are found in many diverse medicinal plant species (e.g., Dong et al., 2021). Triterpene saponins have a wide range of pharmacological applications as anti-cancer, anti-inflammatory, anti-allergic, and anti-viral agents and have value for the treatment of leukemia and hypoglycemia, as well as for the prevention and treatment of cardiovascular and cerebrovascular diseases (Fleck et al., 2006; Ma et al., 2007; Ponou et al., 2008; Sparg et al., 2004; Sun et al., 2004, 2009; Vermeersch et al., 2009; Yan et al., 2006). Among saponins in *A. kteniophylla*, one has shown significant inhibitory effects on six different lines of tumor cells (Gu et al., 2014). Broadly, the *Ardisia* genus appears to represent a rich source of triterpene saponins (Kobayashi & De Mejía, 2005; Su et al., 2003; Zhang, 1994).

In general, the biosynthesis of saponins is well-characterized and primarily involves three major classes of enzymes. These are *oxidosqualene cyclases* (OSCs), *uridin diphosphate glycosyltransferases* (UGTs), and *cytochrome P450 monooxygenases* (CYP450s; Sawai & Saito, 2011). Respectively, these classes of enzymes construct the basic triterpenoid structures, or skeletons, facilitate oxidation reactions, and catalyze the attachment of carbohydrates to a hydroxyl or similar functional groups of other molecules (i.e., glycosylation; Sawai & Saito, 2011).

In plants, saponins are part of a chemically diverse array of secondary metabolites that function in defense against microbes, diseases, and other pests (Papadopoulou et al., 1999). Plant defenses are stimulated through environmental interactions that induce signal transduction, which is often carried out by plant hormones. Among plant hormones, jasmonic acid, its methyl ester, methyl jasmonate (MeJA), and salicylic acid (SA) are endogenous and involved in signal transduction related to defense and, specifically, the production of secondary metabolites (Wang et al., 2015). When SA and MeJA are applied exogenously experimentally, they have been shown to induce upregulation of genes involved in triterpenoid saponin biosynthesis (Cao et al., 2015; Chang et al., 2016).

Unfortunately, the knowledge of genetic and hormonal regulation of biosynthesis of triterpenoid saponins within many medicinal plants, such as *A. kteniophylla*, is constrained by lack of genetic resources, including functionally characterized gene sequences. Characterization and new discoveries of functional genes can be expedited for medicinal plants (and other non-model plant species) using transcriptome sequencing and de novo assembly (Liu et al., 2018; Minoche et al., 2011; Sangwan et al., 2008).

In this study, we sought to elucidate the genomic basis for the biosynthesis of triterpenoid saponins in *A. kteniophylla*. Specifically, our objectives were to (1) detect and characterize genes involved in biosynthesis of triterpenoid saponins using transcriptomic data resulting from RNA sequencing (RNA-Seq), (2) compare the expression levels of genes involved in triterpenoid saponin biosynthesis in leaves and roots of *A. kteniophylla* to better understand the genomic mechanisms underlying disparity in accumulation of these compounds between these two organ systems, and (3) investigate differential levels of transcription of triterpenoid saponin biosynthesis under exposure to the plant hormones, SA and MeJA. Additionally, we assessed the transcription factor families represented by the transcriptomes that we sequenced from roots and leaves to infer possible mechanisms of upstream regulation of the biosynthesis of triterpene saponins in *A. kteniophylla*.

2 | MATERIALS AND METHODS

2.1 | Plant materials

We obtained seeds of *A. kteniophylla* from a group of plants of similar age and height from a nursery in Nanxiong, Guangdong, China. We grew the seeds on mixed soil (coconut bran: perlite:

peat soil = 1:1:1) in a greenhouse at South China Botanical Garden, Guangzhou, Guangdong, China at 20–22°C, which is typical of the average annual temperature in the natural habitat of the species. Based on the preference of *A. kteniophylla* for dark, humid environments (Wei, 2018), we set the greenhouse conditions at 70%–80% humidity and 10%–20% ambient occlusion at all hours of the day. For optimal growth conditions, we sprayed the plants with water three times daily.

After two years of growth, we divided the seedlings of *A. kteniophylla* into SA and MeJA treatment groups and a control group (CK) with six plants in each group. We applied 1 mmol/L concentration of SA or MeJA to leaves of seedlings in the respective treatment groups by spraying until droplets on the leaves were dripping off, and we sprayed leaves of the CK plants with water based on protocols suggested in Werner and Schmülling (2009). All treatments occurred at 8 a.m., 12 a.m., and 6 p.m. daily for a total of seven days based on protocols successfully used for *Ardisia crenata* in a prior study (Yang, 2015).

We selected the concentrations of SA and MeJA and timings of measurements based on prior studies of the medicinal plants, especially *Panax ginseng*, *Ardisia crenata*, and *Astragalus mongholicus* (Cao et al., 2015; Chang et al., 2016; Yang, 2015). For example, the study on *Panax ginseng* (Cao et al., 2015) examined the content of medicinally important ginsenosides and found that 1 mmol/L concentration of SA was effective for their yield and quality. Similarly, a study on *Astragalus membranaceus* found that 1 mmol/L of MeJA was optimal for the species to yield saponins (Chang et al., 2016). Based on these prior findings, we performed a preliminary experiment in which we treated six seedlings of *A. kteniophylla* with 1 mmol/L of MeJA or SA and assessed triterpenoid saponin yield. The preliminary results supported the suitability of application of 1 mmol/L for our experiments.

Following seven days of treatments, we selected three healthy plants out of six from the CK and each treatment group (representing three biological replicates) for sampling of leaves and roots for quantitative real-time PCR (qRT-PCR) and spectrophotometric analysis. We performed both the qRT-PCR and spectrophotometric analyses on the same plants. Additionally, we sampled leaves and roots for RNA extraction and subsequent Illumina sequencing from the same three CK plants. After sampling leaf and root tissues for RNA extraction, we immediately stored the samples in liquid nitrogen and kept them at –80°C until processing.

2.2 | RNA extraction and verification of differences in triterpene saponin concentrations in leaves and roots

We used spectrophotometry to verify that there was a difference in triterpenoid saponin concentrations between leaves and roots in *A. kteniophylla* sufficient to merit further downstream analyses. To prepare tissues for spectrophotometry, we first homogenized 0.1 g

of each tissue sample, placed 1.0 ml of homogenate into 20 ml test tubes, and dried the samples using compressed air. We also generated a control comprising 1.0 ml of distilled water in a 20 ml tube. After the samples were dry, we performed a vanillin-perchloric acid assay (Hiai et al., 1976) to detect the presence of triterpenoid saponins. Specifically, we added 0.2 ml of 5.0% vanillin-glacial acetic acid solution and 1.0 ml of perchloric acid to the tissue and vortexed to mix. Thereafter, we placed the samples in a water bath at 80°C for ten minutes followed by an ice bath for five minutes before adding 8 ml of glacial acetic acid. We used a Specord210Plus spectrometer (Analytikjena) to measure absorbance of the resulting mixture at 550 nm, representing the absorption peak for oleanolic acid, which has the strongest peak among triterpenoid saponins (Hiai et al., 1976). We calculated the concentration of triterpene saponins using the following equation:

$$\text{Content of triterpenoid saponins (mg/g)} = (C * V) / W,$$

where C is the mass of triterpenoid saponins in mg/ml calculated according to the absorbance and the standard curve, V is the volume of the extract, and W is the mass of the sample.

To evaluate differences in the concentrations of triterpenoid saponins among the CK and two experimental groups, we used SPSS22.0 to perform analyses of variance (ANOVAs), and we performed *t*-tests for assessing the differences within groups. In order to generate a standard curve for comparison with the empirical results, we used graduated amounts of oleanolic acid standard solution (1, 2, 3, 4, and 5 ml) and carried out the same color reaction. Absorbance was measured at 550 nm, and a linear regression was performed on the absorbance value A with concentration C (mg/ml).

For RNA extraction, we used 50 mg of tissue with a Huayueyang Quick RNA Isolation Kit according to the manufacturer's protocol. Following extraction, we assessed the quality, purity, and integrity of the total RNA using electrophoresis on a 1.0% (w/V) gel, a NanoPhotometer spectrophotometer, and an Agilent 2100 bioanalyzer. We required that extracted RNAs

TABLE 1 Databases, software, and parameters used in gene annotation

Database	Software and non-default parameters
Nr	Diamond v0.8.22 e-value = 1e–5
Nt	NCBI blast 2.2.28+ e-value = 1e–5
Pfam	HMMER 3.0 package, hmmscan e-value = 0.01
KOG	Diamond v0.8.22 e-value = 1e–3
Swiss-Prot	Diamond v0.8.22 e-value = 1e–5
KEGG	KEGG Automatic Annotation Server e-value = 1e–10
GO	Based on NCBI nr and Pfam results: Blast2GO v2.5 e-value = 1e–6

comprised a concentration of at least 200 ng/μl for downstream processing.

2.3 | Illumina sequencing

We prepared complementary DNA (cDNA) libraries from the extracted RNAs from roots and leaves of *A. kteniophylla*. We generated the libraries using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB) following the manufacturer's recommendations and attached index adaptors to each sample for downstream identification. The sequencing was performed by Novogene Bioinformatics Technology Co. Ltd., using an Illumina novaseq 6000 and paired-end

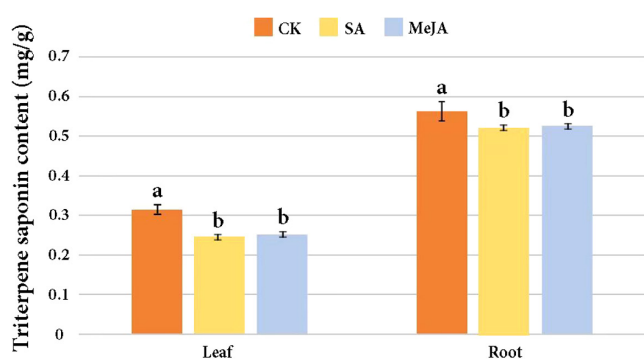


FIGURE 1 Content of triterpene saponins in different experimental and control groups. CK = Control, SA = salicylic acid, MeJA = methyl jasmonate. In each set of three bar charts, samples labeled with the same letter do not show statistically significant differences, while those bearing different letters are statistically different from one another at the $\alpha = .05$ level based on a Tukey's HSD test. Error bars represent standard deviation among three biological replicates

sequencing technology with a read length of 150 bp. The sequencing workflow utilized an Illumina cBot Cluster Generation System with a TruSeq PE Cluster Kit v3-cBot-HS. We submitted the resulting raw reads to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (accession number PRJNA675388).

2.4 | Transcriptome assembly and annotation

We processed the raw reads and obtained high-quality clean sequences using Illumina Casava 1.8 (Richter & Sexton, 2009). Specifically, we removed the adapter sequences as well as reads with ambiguity (i.e., "Ns") and/or Phred scores of ≤ 20 for more than 50% of bases. We performed de novo assembly of the clean, high-quality reads using Trinity (Grabherr et al., 2011) under default settings. Trinity comprises several steps, including a finishing step, "Butterfly," which determines the relative number of reads supporting each path along a de Bruijn graph to disentangle paralogs that were considered to represent a single gene in prior steps (Grabherr et al., 2011). In the earliest version of Trinity, this approached successfully disentangled all paralogs from 43% of multi-gene families and distinguished two or more paralogs from 85% of multi-gene families (Grabherr et al., 2011). Trinity is also known to outperform other assembly algorithms when paralogy is expected due to polyploidy (Chopra et al., 2014). Thus, while we cannot rule out (and fully expect) merger, or collapse, of some closely related paralogs within multi-gene families in our assemblies, we believe that Trinity represents one of the most robust tools available for de novo assembly at present.

From the assembly, we annotated gene functions based on databases comprising NCBI non-redundant protein (nr) and nucleotide sequences (nt), Pfam, EuKaryotic Orthologous Groups

TABLE 2 Summary statistics of transcriptome sequencing data for leaves and roots of *Ardisia kteniophylla*

Sample	Clean reads	Clean bases (Gb)	Error rate	Q20 percentage	Q30 percentage	GC percentage
Sequencing						
root1	22,666,075	6.8	0.03	97.6	93.14	45.81
root2	22,583,212	6.77	0.03	97.17	92.14	45.16
root3	21,714,893	6.51	0.03	97.41	92.68	45.67
leaf1	22,060,727	6.62	0.03	97.42	92.7	45
leaf2	22,079,381	6.62	0.03	97.26	92.3	44.91
leaf3	23,376,880	7.01	0.03	97.45	92.69	44.42
Assembly						
	Transcripts	Unigenes				
Total number	132,318	52,454				
Mean length (bp)	1804	1457				
N50 (bp)	2719	2464				
N90 (bp)	861	572				
Total nucleotides	238,692,243	76,400,059				

(KOG), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO; Table 1) primarily using Blastx, Blastn, or database-specific tools (Conesa et al., 2005). We identified transcription factor (TF) families using iTAK (Zheng et al., 2016), which applies hmmscan (Bilmes, 2006) to compare gene sequences to a profile comprising a set of rules (Jin et al., 2014; Pérez-Rodríguez et al., 2010), in this case, about allowed and disallowed domains in plant TFs based on annotated sequences from several databases.

TABLE 3 Number and percentages of 52,454 unigenes successfully annotated using each sequence database

Database	Number of unigenes	Percentage (%)
NCBI non-redundant protein (nr)	27,400	52.23
NCBI nucleotide sequences (nt)	20,773	39.6
KEGG orthology	9743	18.57
Swiss-Prot	22,231	42.38
Pfam	21,405	40.8
Gene Ontology (GO)	21,405	40.8
KOG	7438	14.18
Annotated by all databases	5808	11.07
Annotated by at least one database	52,454	100

2.5 | Analysis of differentially expressed genes based on transcriptomes and qRT-PCR

We used RSEM (Li & Dewey, 2011) to align the clean reads of each sample to the reference assembly from Trinity and determined the number of reads, or read count, aligned to each gene. We used the average of read counts from the three samples of leaves and roots to infer Foldchange.

We determined differentially expressed genes (DEGs) between the leaves and roots of three samples of *A. kteniophylla* representing the CK group using the DESeq package (1.10.1) in R (Love et al., 2014). DESeq infers DEGs from reads based on read count using a negative binomial distribution with mean and variance inferred via local regression. From the outcome of DESeq, we adjusted the resulting *p*-values using the Benjamini and Hochberg's correction (Benjamini & Hochberg, 2000). We regarded genes found by DESeq with an adjusted log₂Foldchange > 1 and *p*-value < .05 as reliable DEGs between the two types of tissues. To visualize and compare foldchange, we used the R package, DEGseq R package (Wang, Feng, et al., 2010; Wang, Luan, et al., 2010), to generate a volcano plot (Li, 2012), which, broadly, is a scatter plot showing the magnitude of change (*x*-axis) versus statistical significance (*y*-axis).

We also compared DEGs within upstream and downstream parts of the triterpenoid saponin biosynthesis pathway between

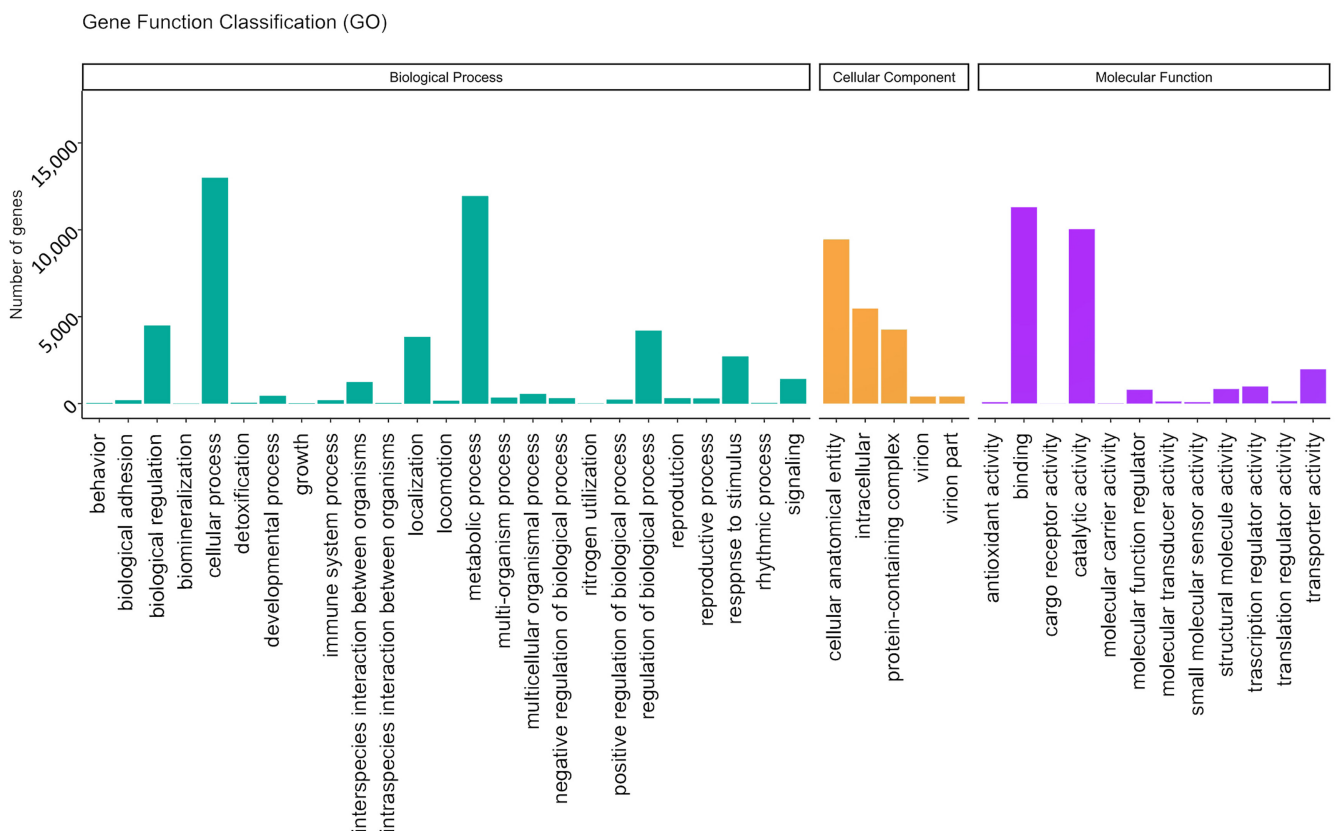


FIGURE 2 Gene Function Classification using GO. Classifications one rank down from the top-level GO categories of Biological Process, Cellular Component, and Molecular Function are shown

leaves and roots. For this purpose, we regarded reactions leading to the production of squalene and 2,3-oxidosqualene from isoprene units of mevalonate (MVA) or methylerythritol phosphate (MEP) as upstream and the production of triterpene saponins from squalene and 2,3-oxidosqualene as downstream. The production of triterpene saponins from squalene and 2,3-oxidosqualene comprises three steps that are catalyzed by *oxidosqualene cyclases* (OSCs), *cytochrome P450 monooxygenases* (CYP450s), and *uridin diphosphate glycosyltransferases* (UGTs), which we anticipated finding among DEGs representing the downstream part of the pathway.

For comparing among root and leaf tissues of the CK and experimental groups, we used quantitative reverse transcriptase PCR (qRT-PCR). We also used qRT-PCR to verify results based on the tally of transcripts for the CK and roots of the experimental groups. For DEGs in leaves, we investigated expression levels using qRT-PCR because the high quality of extracted RNAs from leaf tissues made this approach the most reliable.

For qRT-PCR, we reverse-transcribed total RNA with a GoScript™ Reverse Transcription Mix (Promega) according to the manufacturer's protocol and performed reactions using a SYBR® qPCR Master Mix on a Roche LightCycler480II (Roche) with gene-specific primer pairs. Primers were developed for this study based on the coding sequences of the genes (Table S1). As a standard, we used *Beta-Actin* (β -ACTIN; a "housekeeping" gene) and calculated its relative expression between the two types of tissue using $2^{-\Delta\Delta C_t}$ (Livak & Schmittgen, 2001).

3 | RESULTS

3.1 | Determination of the content of triterpene saponins

Within both treatment groups and the CK, the content of triterpene saponins in root tissues of *A. kteniophylla* was much higher than in leaf tissues (Figure 1) based on spectrophotometry. In both the leaves and roots, the CK bore significantly higher concentrations of triterpene saponins than either of the treatment groups ($\alpha = 0.05$). However, we observed only negligible differences in concentrations of triterpene saponins between the two treatment groups for the leaves and roots.

3.2 | Sequencing and sequence assembly

After determining that there were differences in concentrations of triterpenoid saponins between the leaves and roots of *A. kteniophylla* using spectrophotometry, we performed Illumina sequencing of these tissues from CK plants representing three biological replicates from three different plants. The resulting sequences comprised an average of 22,505,663 high-quality reads from leaves and 22,321,393 high-quality reads from roots following filtering and 97.38% and 97.39% of sequences, respectively, had Phred quality scores of 20 (Q20) or

better. The *de novo* assembly in Trinity yielded 132,319 transcripts and 52,454 unigenes (File S1) with N50 of 2719 and 2464 bp for the leaves and roots, respectively, and the mean length of transcripts and genes was 1804 and 1457 bp, respectively (Table 2).

3.3 | Functional annotation and classification

We were able to annotate all 52,454 unigenes using at least one source among the NCBI non-redundant protein (nr) and nucleotide sequences (nt) databases, Pfam, KOG, Swiss-Prot, KEGG, and the Gene Ontology (GO) database (Table 3). This represents a 100% success rate in annotation of unigenes. In the case of 5808 unigenes, we performed successful annotation using all databases.

In GO, the largest proportion of genes in the biological process category were annotated as "cellular process" (GO:0009987;

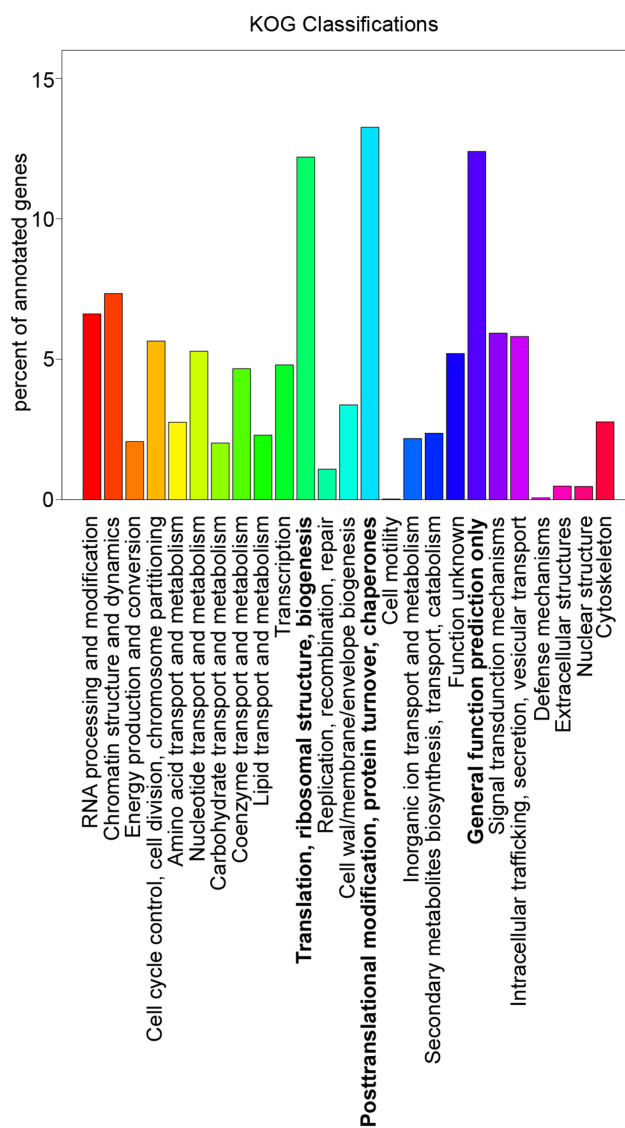


FIGURE 3 Gene Function Classification using KOG. The three classifications showing the highest enrichment are shown in bold type

KEGG Classification

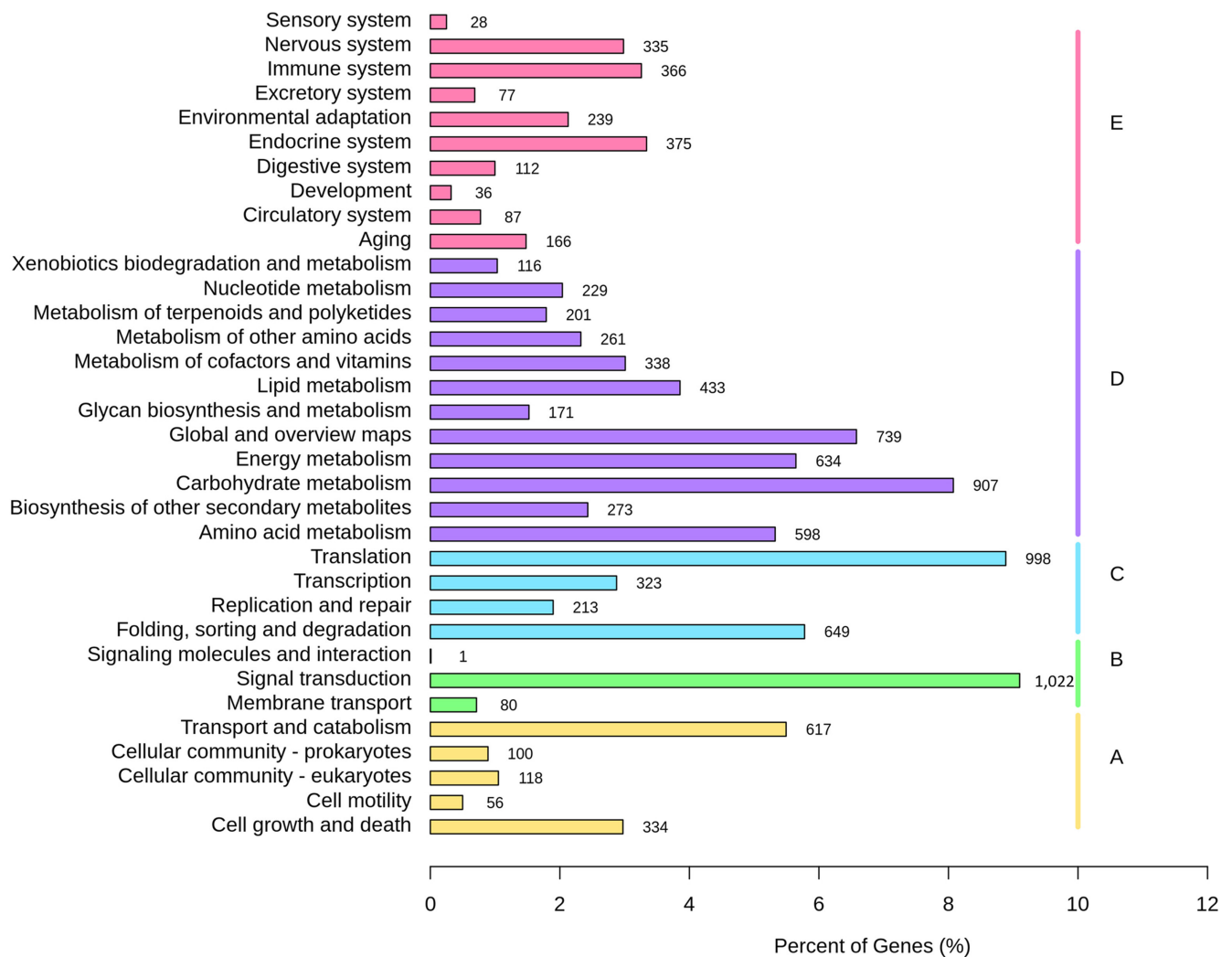


FIGURE 4 Kyoto Encyclopedia of Genes and Genomes classification of 9743 assembled unigenes. Categories of annotations to pathways shown to the right comprise cellular processes (A), environmental information processing (B), genetic information processing (C), metabolism (D), and organismal systems (E). For each pathway, the number of genes is shown to the right of the bar, which corresponds to the total percentage of genes indicated by the x-axis

12,989 genes, 28.26%), within the cellular component category, the largest proportion were annotated as “cellular anatomical entity” (GO:0110165; 9457 genes, 47.20%), while “binding” (GO:0005488; 11,316 genes, 42.68%) was the most common annotation within the molecular function category (Figure 2).

In KOG, we annotated 7438 genes according to their functional categories (Figure 3). The top three categories represented by the genes were “translation, ribosomal structure and biogenesis” (907 annotated genes, 12.20%), “posttranslational modification, protein turnover, chaperones” (986 annotated genes, 13.25%), and “general function prediction only” [usually representing biochemical activity (Tatusov, 2001) 922 annotated genes, 12.40%].

We used the KEGG database to map 9743 genes of *A. kteniophylla* to five total biochemical pathways (Figure 3). These consisted of cellular processes, environmental information processing,

genetic information processing, metabolism, and organismal systems (Figure 4). For each biochemical pathway, the secondary pathways with the highest proportion of annotated genes were “transport and catabolism” (617 genes, 6.33%), “signal transduction” (1022 genes, 10.49%), “translation” (998 genes, 10.24%), “carbohydrate metabolism” (907 genes, 9.31%), and “endocrine system” (375 genes, 3.85%), respectively.

3.4 | Differential expression analysis

We identified DEGs between leaves and roots of *A. kteniophylla* in DESeq (Table S2) and detected 6092 unigenes upregulated in the roots and 6001 unigenes upregulated in leaves (Figure 5). Among DEGs, 7065 showed significant differential expression, and these

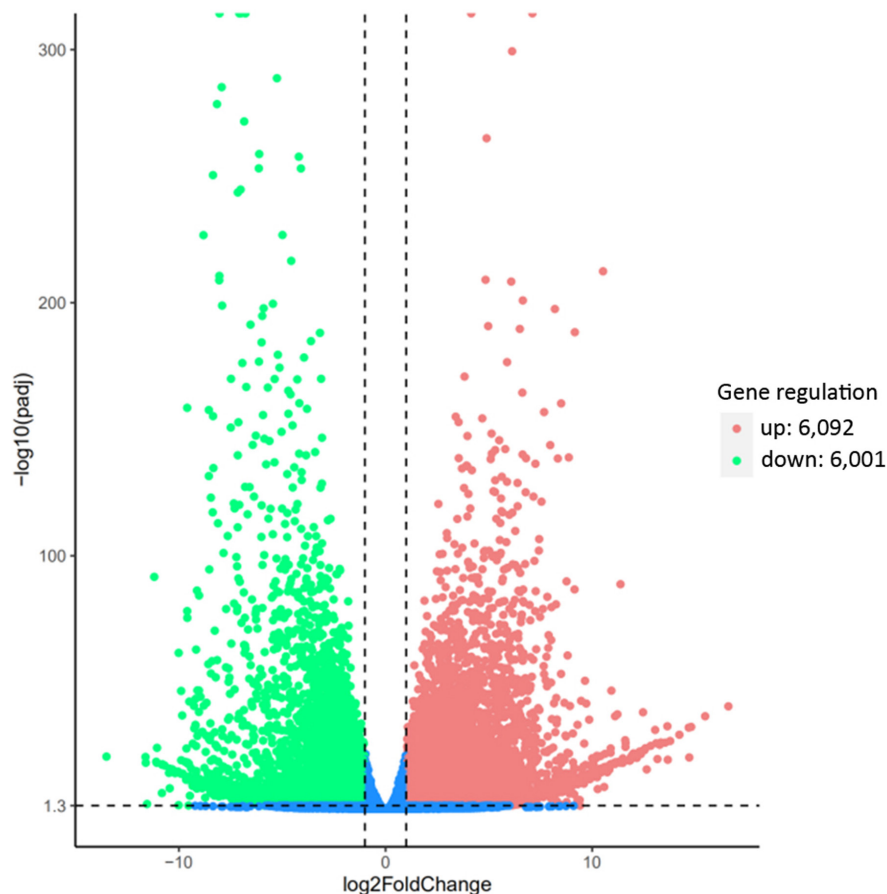


FIGURE 5 Volcano plot showing differentially expressed genes in roots of *Ardisia kteniophylla* compared with leaves. The x-axis indicates the fold change in expression for different genes, and the y-axis indicates the significance level of the difference in expression. The blue dashed line represents the threshold for determining that differential expression is significant

represented 4307 GO terms. Among the GO terms, 2495, 606, and 1215 DEGs were involved in biological processes, comprised cellular components, and were related to molecular functions, respectively. Within the biological process category, the greatest portion of DEGs were mapped to “protein phosphorylation” (408 DEGs, 16.35%; GO: 0006468). For DEGs associated with the cellular component category, the majority were mapped to “tubulin complex” (76 DEGs, 12.54%; GO: 0045298), and DEGs representing molecular function were most frequently involved in “oxidoreductase activity” (852 DEGs, 70.12%; GO: 0016705; [Figure 6](#)).

We mapped 2737 DEGs onto 117 pathways in the KEGG database. The top three pathways involving DEGs were “plant hormone signal transduction” (108, 3.95%; KO:04075), “starch and sucrose metabolism” (103, 3.76%; KO:00500), and “phenylpropanoid biosynthesis” (98, 3.58%; KO:00940). Moreover, the KEGG pathways “phenylpropanoid biosynthesis,” “flavonoid biosynthesis,” and “photosynthesis” contained DEGs showing the greatest differences between the root and leaves of *A. kteniophylla* based on *q* values (i.e., a type of corrected *p*-values; [Figure 7](#)).

3.5 | Genes in the up- and downstream pathways of triterpene saponin biosynthesis

We found that DEGs involved in the upstream MVA pathway, such as *3-hydroxy-3-methylglutaryl-CoA reductase* (*AkHMGR*), *5-diphosphomevalonate decarboxylase* (*AkPMD*), were more

abundant in roots compared with the leaves. In contrast, DEGs associated with the MEP pathway, which showed greater abundance of genes in leaves, included *2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase* (*AkMDS*) and *4-hydroxy-3-methylbut-2-enyldiphosphate synthase* (*AkHDS*) ([Table S3](#)). Based on \log_2 (fold change), we found that genes involved in the final steps of upstream biosynthesis processes, such as *squalene synthase* (*AkSS*) and *squalene monooxygenase* (*AkSM*), were more highly expressed in roots, while leaves showed greater expression of genes involved in biosynthesis of earlier precursors ([Figure 8](#)).

From among DEGs in the downstream part of the triterpene saponin biosynthesis pathway, we identified two *OSCs*, 144 *CYP450s*, and 31 *UGTs*, of which 0, 100, and 22 satisfied the criterion of \log_2 fold change ≥ 1 , respectively. Overall, more DEGs in the downstream pathway were more highly expressed in roots compared with leaves ([Table S4](#)). None of the predicted *CYP450s* or *UTGs* were annotated to the KEGG pathways for triterpenoid biosynthesis ([Figure 8](#)), and this may be because the functions of these genes are still broadly unknown and are not presently included in the focal pathways by KEGG.

3.6 | Quantitative reverse transcriptase PCR

Based on the DEGs inferred from transcriptome sequences, we selected the following genes for qRT-PCR: two *AkPMD* unigenes, one *AkHDS* unigene, one *AkSS* unigene, two *AkSM* unigenes, five *AkCYP450s* unigenes, and four *AkUGTs* unigenes. We found that all

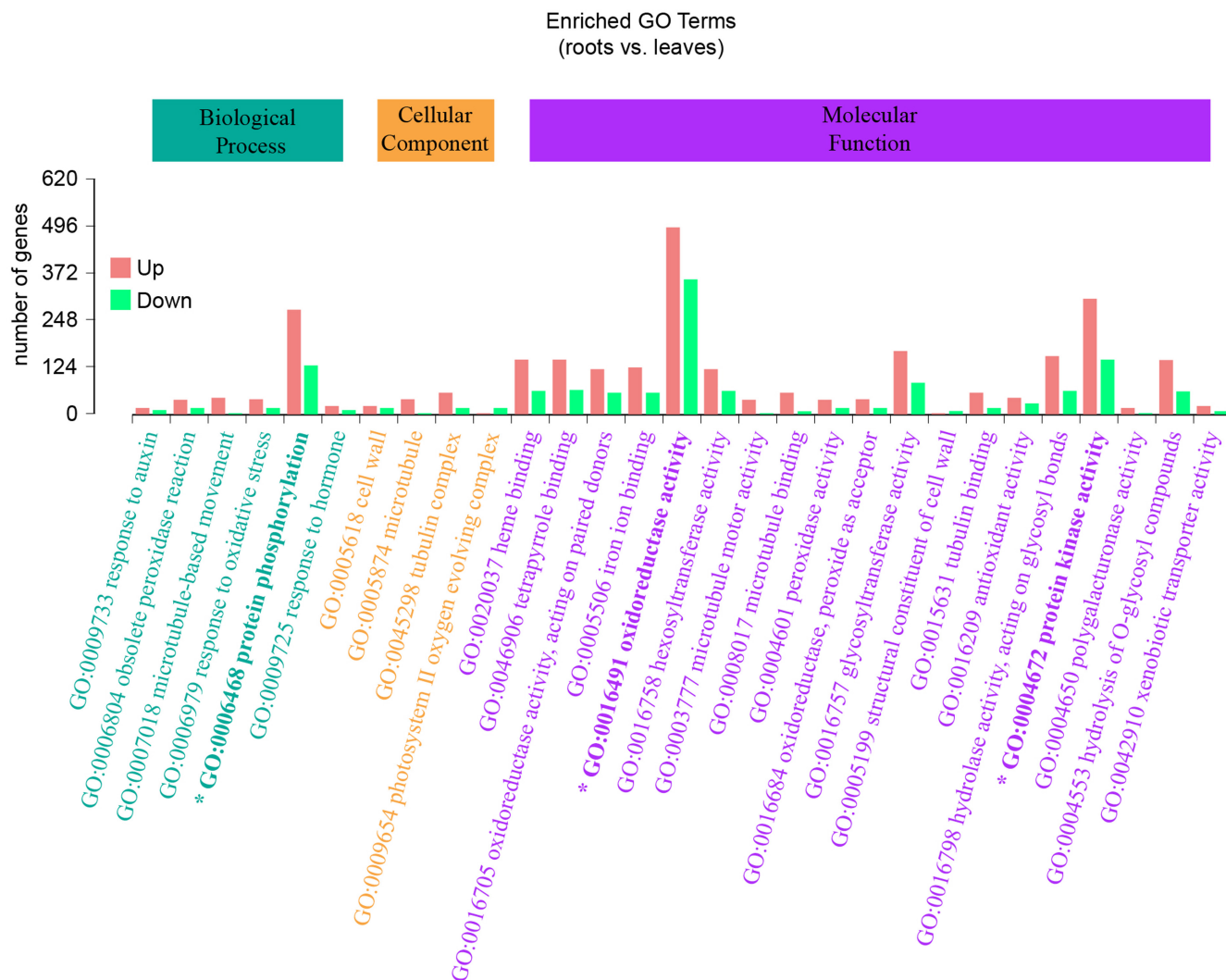


FIGURE 6 Gene Ontology annotations of DEGs in *Ardisia kteniophylla*. Terms shown on the x-axis are one level below the highest-level terms: biological process (BP), cellular component (CC), and molecular function (MF). Bold type and asterisks indicate the three terms with the largest number of DEGs overall

genes analyzed using qRT-PCR were upregulated in the roots of the CK plants (Figure 9) as is generally consistent with the DEGs inferred from the transcriptomes (Table S1).

Compared with the CK plants, genes operating upstream within the triterpene saponin biosynthesis pathway (*AkPMDs*, *AkHDS*, *AkSMs*, and *AkSS*) were upregulated in both the SA and MeJA treatment groups (Figure 10). In contrast, all genes involved in the downstream pathway (*AkCYP450s* and *AkUGTs*) showed downregulation in the treatment groups compared with the CK except *AkCYP3*, which was upregulated under SA treatment (Figure 10i). Overall, regulation of genes in the up- and downstream pathways showed greater up- or downregulation in response to SA than to MeJA.

3.7 | Prediction of transcription factors

In total, we annotated 997 transcription factors in *A. kteniophylla*, and, of these, 590 were upregulated in the root and 401 were upregulated in leaves (Table 4 and Table S5). In particular, members

of the *Myeloblastosis* (*MYB*) and *Apetala 2/Ethylene Response Factor* (*AP2/ERF*) TF families were the most highly upregulated in roots with 95 and 76 members, respectively. Notably, *MYB* was also the most highly upregulated in leaves with 65 upregulated members. Overall, we found that the 1865 potential TFs accounted for 3.56% of the *A. kteniophylla* unigenes library.

4 | DISCUSSION

4.1 | Triterpenoid saponin biosynthesis genes of *A. kteniophylla* and their expression levels in leaves and roots

In *A. kteniophylla*, we detected a larger number of unigenes over 1 kb in length compared with its congener, *Ardisia crenata* (Yang, 2015); 23,568 versus 14,659, respectively. The difference in our results might be due to the availability of more data within public databases in contrast to six years ago when the prior study was published and

Statistics of Pathway Enrichment

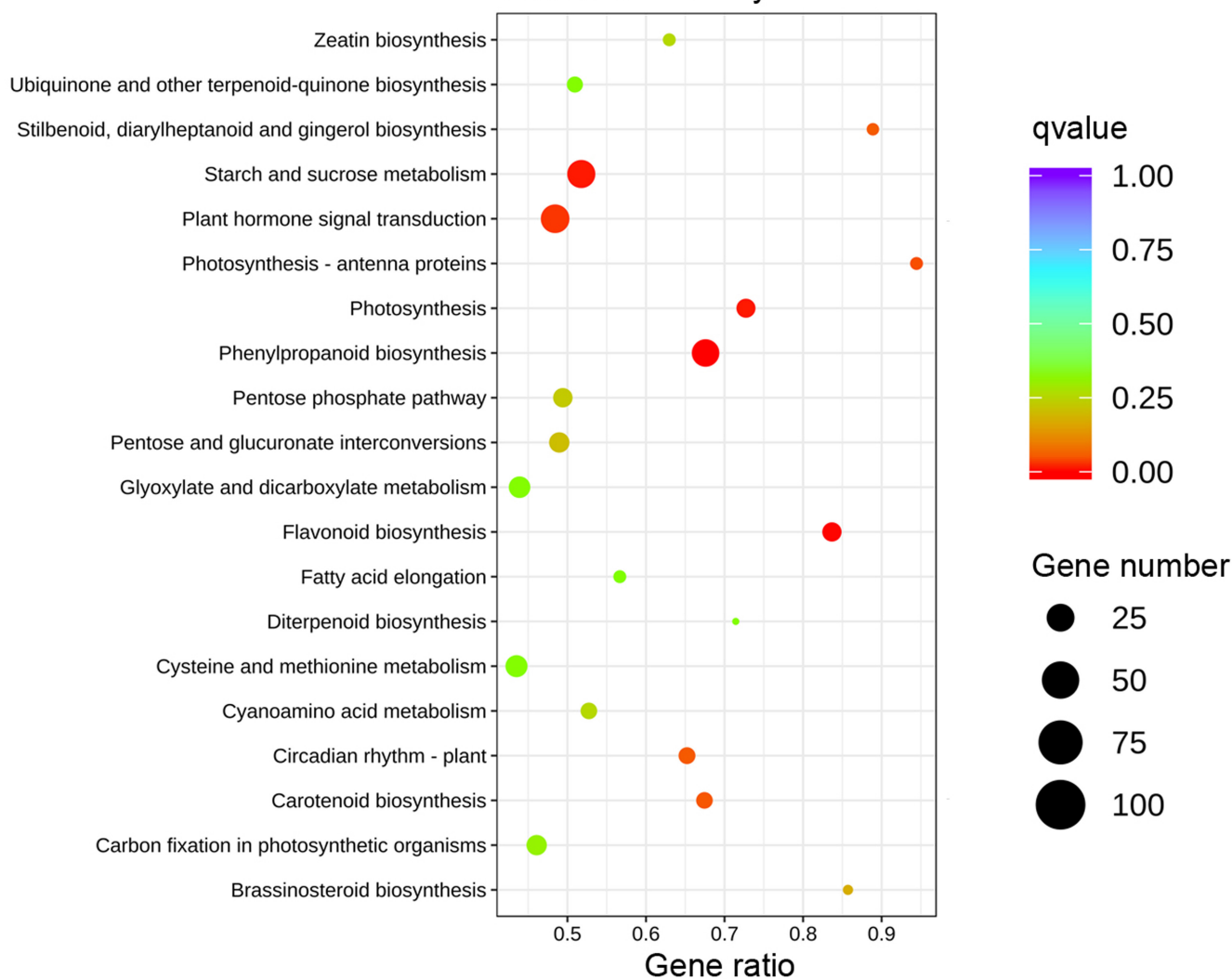


FIGURE 7 Enrichment of DEGs in KEGG pathways. The number of DEGs contained in each pathway is represented by the size of the dot

to improvements in transcriptome sequencing capabilities (Stark et al., 2019). The long lengths of unigenes generated in this study suggest an overall high-quality transcriptome (Wang, Feng, et al., 2010; Wang et al., 2012; Wang, Luan, et al., 2010) that is likely responsible for our 100% success rate at annotating all 52,454 unigenes that we recovered.

We detected 2737 DEGs between leaves and roots of *A. kteniophylla* and mapped these to 117 pathways in the KEGG database (Figure 7). Of the 2737 DEGs, 108 (3.95%) were mapped to the “plant hormone signal transduction” pathway (KO:04075), and this was the largest number of genes mapped to any one pathway. Notably, genes participating in signal transduction may induce the expression of key enzymes that are involved in the production of secondary metabolites (Arimura et al., 2000), such as triterpene saponins (Yendo et al., 2014). Several specific signal transduction pathways influencing the production of triterpene saponins have been studied in medicinal species of *Panax*, and signaling mechanisms in those species include calcium, ethylene, and nitric oxide, and reactive oxygen species

(Rahimi et al., 2015). However, more research is needed to directly connect the signal transduction enzymes that we detected in the roots and leaves of *A. kteniophylla* to levels of accumulation of triterpene saponins.

Our results show that the two pathways for generation of 2,3-oxidosqualene from isoprene units occur differentially in the leaves and roots based on DEGs. Within the roots, the MVA pathway is more active in the generation of 2,3-oxidosqualene from isoprene units, while, in leaves, the MEP pathway is more active. These findings were consistent with previous studies, which have shown that genes representing the MVA pathway are most highly expressed in the radicle, hypocotyl, roots, flowers, and seeds, whereas genes of the MEP pathway are more active in leaves (Vranová et al., 2013). This is likely because the MEP pathway is confined to plastids (Lichtenthaler, 1999; Rodríguez-Concepción et al., 2004), which are usually most abundant in the leaves as chloroplasts. Moreover, while the MEP pathway may be technically operational within plastids that have not been exposed to light, that is, etioplasts, precursory components for initiating the

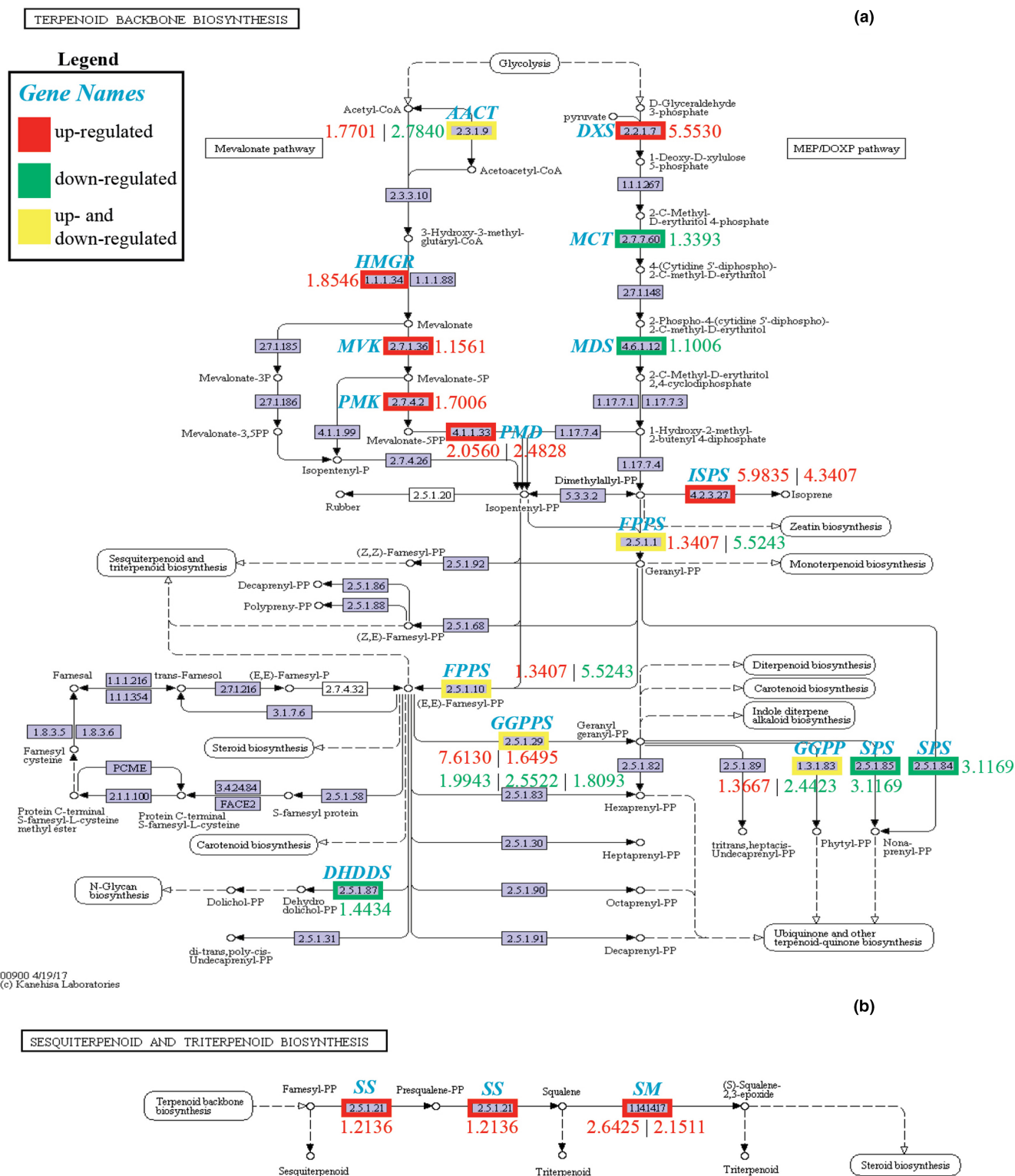


FIGURE 8 Enrichment of DEGs for the KEGG pathways. (a) terpenoid backbone biosynthesis pathway (ko00900) and (b) sesquiterpenoid and triterpenoid biosynthesis (ko00909), which are both related to triterpene saponins synthesis. Numbers below or adjacent to nodes indicate the ratio of \log_2 (fold change). Genes are up or downregulated in roots compared with leaves

MEP pathway appear to be lacking in dark environments, thus limiting the utility of this pathway in roots (Nagata et al., 2002).

Differentially expressed genes in the downstream part of the triterpene saponin biosynthesis were, overall, more highly

expressed in roots compared with leaves (Table S4). This is consistent with higher concentrations of triterpene saponins within the roots of *A. kteniophylla* based on prior studies using biochemical assays (Wei, 2017) and our own results (Figure 1). Moreover, prior

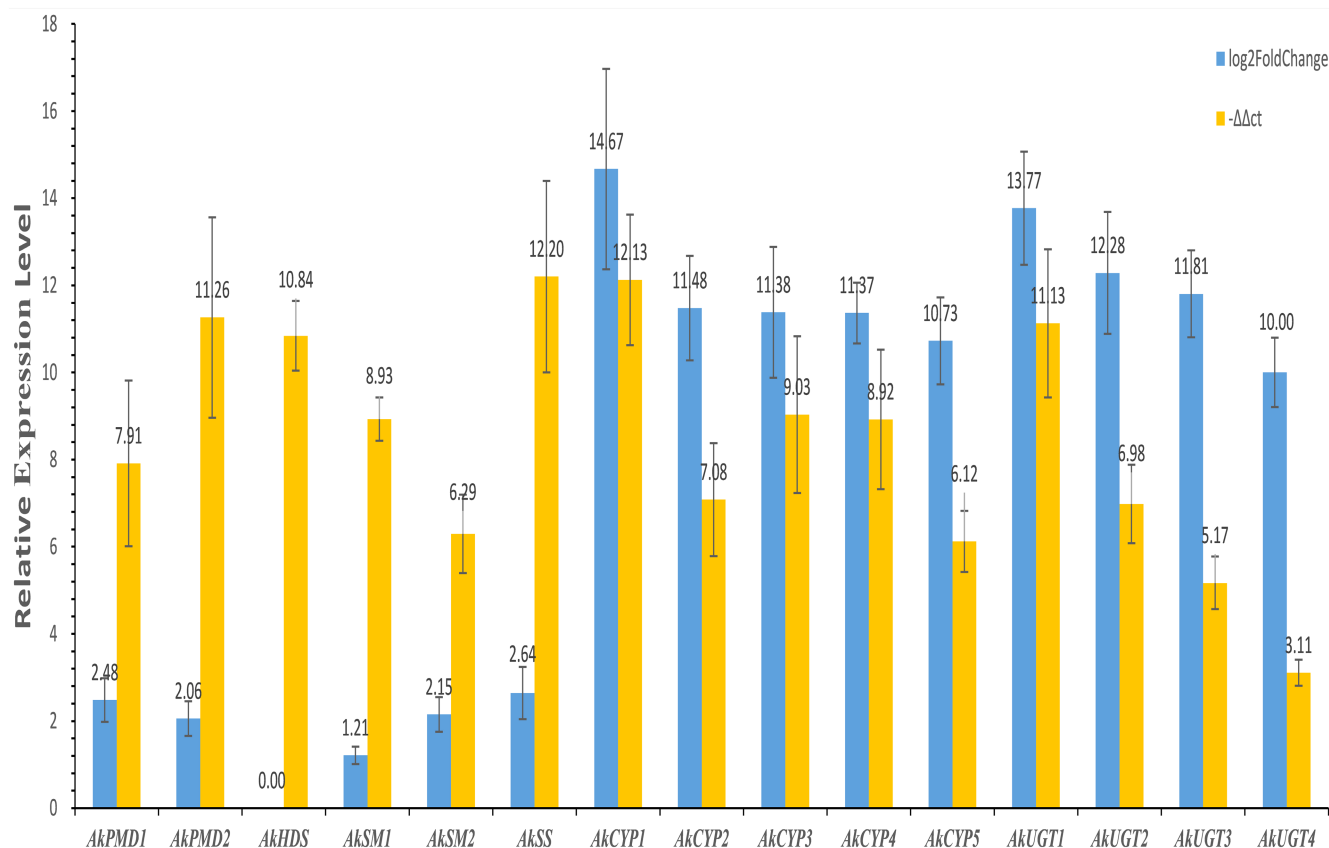


FIGURE 9 Relative expression levels of DEGs selected for qRT-PCR and transcriptome in roots relative to leaves of *Ardisia kteniophylla*. Comparison of log2Foldchange in the transcriptome data with the $-\Delta\Delta C_t$ in the qRT-PCR experiment. Error bars represent the standard deviation among three biological and three technical replicates

studies using gene expression analysis have often shown higher levels of gene expression in the roots of medicinal plants compared with leaves (e.g., *Medicago truncatula*, Fabaceae; Huhman et al., 2005), even in cases where the leaf is the organ preferred for pharmacology (e.g., *Hedera helix*, Araliaceae; Sun, Li, et al., 2017; Sun, Jiang, et al., 2017).

Notably, roots may often be the final site of biosynthesis and storage of triterpene saponins (Sawai and Saito, 2011), while earlier, upstream stages of synthesis occur in leaves (and other tissues not sampled in this study, such as stems, rhizomes, and flowers) as shown in the medicinal plants, *Platycodon grandiflorum* (Campanulaceae; Ma et al., 2016) and *Panax japonicus* (Araliaceae; Rai et al., 2016). In the medicinal herb, *Panax ginseng*, CYP450s are more highly expressed in rhizomes, while UGTs are more highly expressed in roots, suggesting these organs may be, respectively, the intermediate and final sites of triterpene saponin biosynthesis (Han et al., 2012; Rai et al., 2016). Overall, plant roots have often been shown to accumulate higher concentrations of secondary metabolites, such as triterpenoid saponins, and this may be due to the complex nature of the soil environment, where roots are forced into close contact with potentially harmful microbes and pests, thus necessitating strong defenses (Su et al., 2005).

4.2 | Expression of triterpenoid saponin biosynthesis genes of *A. kteniophylla* under exposure to SA and MeJA plant hormones

We found that exogenous application of the signal transduction hormones, SA and MeJA, yielded an overall decrease in the content of triterpene saponins compared with the CK (Figure 1). Research in other plant groups has shown that exogenous application of MeJA can increase the production of triterpenoid saponins compared with control plants, such as in *Centella asiatica* (Apiaceae; Mangas et al., 2006). Similar results were found in *Nigella sativa* (Ranunculaceae) under exogenous application of SA (Elyasi et al., 2016). However, both hormones are known to limit plant growth, potentially resulting in negative interactions between growth-related pathways and production of secondary compounds (Ellis & Turner, 2001; Rudell et al., 2002). The degree to which plants are able to both increase secondary compound production while plant growth is stunted under SA or MeJA is likely taxon-specific. For example, in *Ginkgo biloba*, the production of terpene trilactones was increased under SA application even while genes involved in photosynthesis (and, thus growth) were downregulated (Ye et al., 2020), and this is in contrast to our

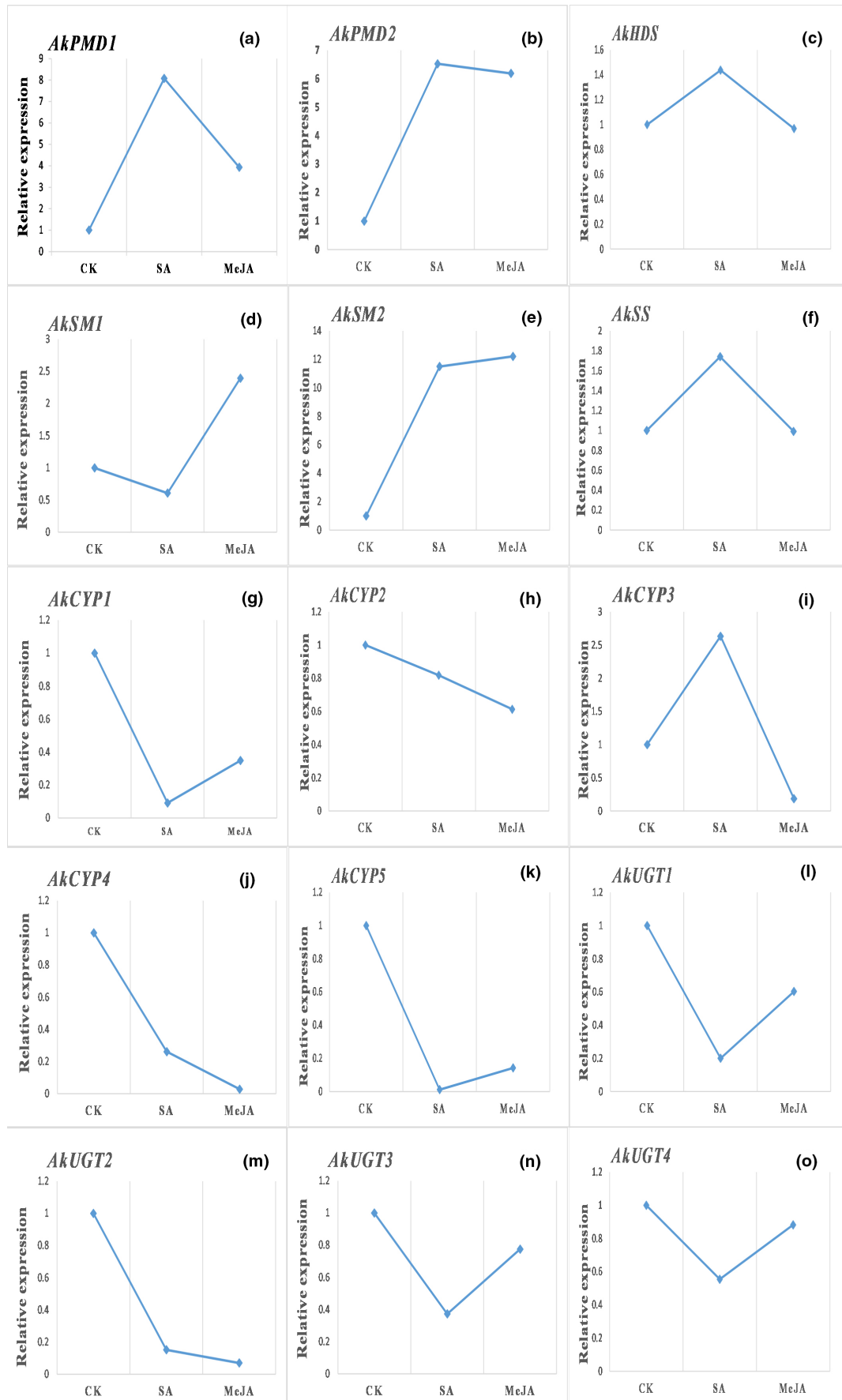


FIGURE 10 Relative expression levels of DEGs selected for qRT-PCR in hormone treatment group and control group. For comparisons, we set the relative abundance of each gene in control group to 1.0. Each point represents the average of three biological replicates

TABLE 4 Differentially expressed transcription factor families in *Ardisia kteniophylla*

Transcription factor family	Upregulated in root	Upregulated in leaf	Total detected
AP2/ERF	76	37	114
NAC	54	37	91
C2H2	50	29	80
MYB	95	65	161
bHLH	51	27	79
WRKY	36	33	69
bZIP	29	23	52
C3H	22	34	56
GRAS	26	17	43
SNF2	24	17	41
B3	21	20	41
GARP	22	17	39
LOB	18	9	29
SET	24	13	37
Trihelix	21	8	29
PHD	21	15	36

results. In taxa where there is stunting of plant growth, this may represent a trade-off where survival in the face of pests is given precedence.

In the case of *A. kteniophylla*, application of SA and MeJA appeared to upregulate genes involved in the early stages of biosynthesis of triterpene saponins compared with the CK, but the critical downstream gene families, *P450* and *UGT*, were downregulated. There is a known antagonistic relationship between *P450* and the effects of MeJA on the production of secondary metabolites involved in defense in *Arabidopsis* (Lee et al., 2011), which can be further explored in future studies within *A. kteniophylla*. Overall, exogenous application of SA or MeJA to leaves of *A. kteniophylla* does not improve the production of the pharmacologically valuable triterpenoid saponins in the species, suggesting that more work is needed to understand the complexities of the biosynthesis pathway for these compounds.

4.3 | Transcription factors families that may regulate triterpenoid saponin biosynthesis in *A. kteniophylla*

Our analyses revealed that TFs previously reported present in *A. kteniophylla*, *WRKY*, *AP2/ERF*, *NAC*, *MYB*, and *basic helix-loop-helix* (*bHLH*; De Geyter et al., 2012; Yang et al., 2012), were always more active in roots. For example, 76% of genes in the *AP2/ERF* family that we identified were more highly expressed in roots and similarly with 54% of *NAC* genes. Moreover, the high overall prevalence of TFs in the genome of *A. kteniophylla* that we can infer from our transcriptomes (i.e., 3.56%) is roughly consistent with *Arabidopsis*, in which

1500 TFs comprise ca. 5% of the genome (Riechmann et al., 2000). A slight difference here may be due to a larger number of TFs detected within the whole genome of *Arabidopsis* compared with our utilization of transcriptomes.

Members of the *AP2/ERF* TF family have been shown to regulate secondary metabolic pathways in several medicinal plants. For example, in *Artemisia annua* (Asteraceae), two *AP2/ERF* TFs, *AaERF1* and *AaERF2*, are responsive to jasmonic acid (JA) and positively regulate genes encoding amorpho-4,11-diene synthase and *CYP71AV1*, enzymes directly involved in biosynthesis of the secondary metabolite, artemisinin (Yu et al., 2012), which is widely used medicinally. Similarly, the *AP2/ERF* TFs, *ORCA2* and *ORCA3*, in *Catharanthus roseus* (Apocynaceae) regulate the metabolism of pharmacologically valuable terpenoid indole alkaloids by activating the expression of the strictosidine synthase gene (Menke et al., 1999; Van Der Fits and Memelink, 2001). In *A. kteniophylla*, the large percentage of detected *AP2/ERF* TFs that were more highly expressed in roots may indicate a critical role of this TF family in the downstream stages of triterpenoid saponin biosynthesis for this species. Thus, *AP2/ERF* genes may represent critical targets for molecular breeding of this species with improved concentrations of medicinally active compounds.

5 | CONCLUSIONS

In summary, in this study, we performed transcriptomic analysis to investigate the genetic basis for accumulation of triterpene saponins in roots and leaves of *A. kteniophylla*. We generated a large dataset of unigenes and successfully annotated these for structure and function based on several public databases. Our comparative analyses revealed that many genes involved in triterpene saponin biosynthesis have higher expression in roots than leaves, and roots are the primary organ used pharmacologically in traditional Chinese medicine. Our transcriptomic data provide a valuable genetic resource for this plant and may facilitate molecular breeding to meet demand for the medicinally bioactive triterpene saponins of the species.

AUTHOR CONTRIBUTIONS

Yuyang Lei: Conceptualization (equal); Formal analysis (lead); Investigation (lead); Writing – original draft (lead); Writing – review & editing (equal). **AJ Harris:** Project administration (supporting); Writing – review & editing (equal). **Aihua Wang:** Formal analysis (supporting). **Liyun Zhao:** Formal analysis (supporting); Investigation (supporting); Writing – original draft (supporting). **Ming Luo:** Conceptualization (supporting). **Ji Li:** Investigation (supporting). **Hongfeng Chen:** Conceptualization (equal); Project administration (lead); Writing – review & editing (equal).

ACKNOWLEDGEMENTS

We are grateful to the Key Laboratory of Plant Resources Conservation and Sustainable Utilization of South China Agricultural Plant Molecular Analysis and Genetic Improvement

and the Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement for providing guidance and laboratory equipment for saponin extraction experiments and molecular biology experiments. We also thank Novogene Bioinformatics Technology Co. Ltd. (Beijing, China) for performing the transcriptome sequencing and part of the downstream bioinformatics analysis.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

Illumina short reads are available from NCBI as PRJNA675388. All other raw data are included as supporting information.

ORCID

Yuyang Lei  <https://orcid.org/0000-0002-3456-5160>

AJ Harris  <https://orcid.org/0000-0003-3215-1201>

Aihua Wang  <https://orcid.org/0000-0002-9940-2071>

Liyun Zhao  <https://orcid.org/0000-0002-8618-5944>

Ming Luo  <https://orcid.org/0000-0002-6912-7638>

Hongfeng Chen  <https://orcid.org/0000-0002-8415-3260>

REFERENCES

- Arimura, G.-I., Ozawa, R., Shimoda, T., Nishioka, T., Boland, W., & Takabayashi, J. (2000). Herbivory-induced volatiles elicit defence genes in lima bean leaves. *Nature*, 406, 512–515. <https://doi.org/10.1038/35020072>
- Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25, 60–83. <https://doi.org/10.3102/10769986025001060>
- Bilmes, J. A. (2006). What HMMs can do. *IEICE Transactions on Information and Systems*, 89, 869–891. <https://doi.org/10.1093/ietisy/e89-d.3.869>
- Cao, W. L., Meng, X. C., & Ma, W. (2015). Effects of salicylic acid on photosynthetic characteristics, physiology and biochemistry and saponins content of ginseng under full light in spring. *Chinese Journal of Chinese Materia Medica*, 40(18), 3553–3559.
- Chang, Y., Yan, S., Liu, Z. P., Ren, W. C., Ma, W., & Liu, J. (2016). Study on screening and regulating elicitor of astragaloside content in plant. *China Journal of Traditional Chinese Medicine and Pharmacy*, 31(8), 3227–3230.
- Chopra, R., Burow, G., Farmer, A., Mudge, J., Simpson, C. E., & Burow, M. D. (2014). Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) rna-seq data. *PLoS One*, 9(12), e115055. <https://doi.org/10.1371/journal.pone.0115055>
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21, 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Dai, W. B., Dong, P. P., & Mei, Q. X. (2017). The textual research of *Ardisia gigantifolia* and its confounding products. *Journal of Chinese Medicinal Materials*, 40, 2220–2223.
- De Geyter, N., Gholami, A., Goormachtig, S., & Goossens, A. (2012). Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends in Plant Science*, 17, 349–359. <https://doi.org/10.1016/j.tplants.2012.03.001>
- Dong, S., Liu, M., Liu, Y., Chen, F., Yang, T., Chen, L. U., Zhang, X., Guo, X., Fang, D., Li, L., Deng, T., Yao, Z., Lang, X., Gong, Y., Wu, E., Wang, Y., Shen, Y., Gong, X., Liu, H., & Zhang, S. (2021). The genome of *Magnolia biondii* Pamp. provides insights into the evolution of Magnoliales and biosynthesis of terpenoids. *Horticulture Research*, 8, 1–13. <https://doi.org/10.1038/s41438-021-00471-9>
- Ellis, C., & Turner, J. G. (2001). The *Arabidopsis* mutant *cev1* has constitutively active jasmonate and ethylene signal pathways and enhanced resistance to pathogens. *The Plant Cell*, 13, 1025–1033.
- Elyasi, R., Majdi, M., Bahramnejad, B., & Mirzaghaderi, G. (2016). Spatial modulation and abiotic elicitors responses of the biosynthesis related genes of mono/triterpenes in black cummin (*Nigella sativa*). *Industrial Crops and Products*, 79, 240–247. <https://doi.org/10.1016/j.indcrop.2015.11.005>
- Fleck, J. D., Kauffmann, C., Spilki, F., Lencina, C. L., Roehe, P. M., & Gosmann, G. (2006). Adjuvant activity of *Quillaja brasiliensis* saponins on the immune responses to bovine herpesvirus type 1 in mice. *Vaccine*, 24, 7129–7134. <https://doi.org/10.1016/j.vaccine.2006.06.059>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Gu, Y., Mu, L., Dong, X., Yao, C., & Liu, P. (2014). Effect of triterpenoid saponins H1 from *Ardisia gigantifolia* on proliferation of six different tumor cell lines and apoptosis and cell cycle of A549 tumor. *Chinese Journal of Experimental Traditional Medical Formulae*, 20, 130–133.
- Han, J. Y., Hwang, H. S., Choi, S. W., Kim, H. J., & Choi, Y. E. (2012). Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. *Plant and Cell Physiology*, 53, 1535–1545. <https://doi.org/10.1093/pcp/pcs106>
- Hiai, S., Oura, H., & Nakajima, T. (1976). Color reaction of some saponin and saponins with vanillin and sulfuric acid. *Planta Medica*, 29(2), 116–122. <https://doi.org/10.1055/s-0028-1097639>
- Huang, G. H., Hu, C. M., & Hao, G. (2017). Rediscovery of *Ardisia gigantifolia* and the reinstatement of *A. kteniophylla* (Primulaceae)—finding the correct name for the Chinese medicinal plant 'Zou Ma Tai'[J]. *Nordic Journal of Botany*, 35, 628–632.
- Huhman, D. V., Berhow, M. A., & Sumner, L. W. (2005). Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. *Journal of Agriculture and Food Chemistry*, 53, 1914–1920. <https://doi.org/10.1021/jf0482663>
- Jin, J., Zhang, H., Kong, L., Gao, G., & Luo, J. (2014). PlantTFDB 3.0: A portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research*, 42, D1182–D1187. <https://doi.org/10.1093/nar/gkt1016>
- Kobayashi, H., & De Mejía, E. (2005). The genus *Ardisia*: A novel source of health-promoting compounds and phytopharmaceuticals. *Journal of Ethnopharmacology*, 96, 347–354. <https://doi.org/10.1016/j.jep.2004.09.037>
- Lee, I. C., Hong, S. W., Whang, S. S., Lim, P. O., Nam, H. G., & Koo, J. C. (2011). Age-dependent action of an ABA-inducible receptor kinase, RPK1, as a positive regulator of senescence in *Arabidopsis* leaves. *Plant and Cell Physiology*, 52, 651–662. <https://doi.org/10.1093/pcp/pcr026>
- Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12, 1–16. <https://doi.org/10.1186/1471-2105-12-323>
- Li, W. (2012). Volcano plots in analyzing differential expressions with mRNA microarrays. *Journal of Bioinformatics & Computational*

- Biology, 10(6), 1231003. <https://doi.org/10.1142/S0219720012310038>
- Lichtenthaler, H. K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual Review of Plant Biology*, 50, 47–65. <https://doi.org/10.1146/annurev.arplant.50.1.47>
- Liu, F. X., Yang, W. G., & Sun, Q. H. (2018). Research based on transcriptome sequencing data analysis and high-throughput GO annotation theory. *Journal of Anhui Agricultural Sciences*, 46, 88–91+100.
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*, 25, 402–408. <https://doi.org/10.1006/meth.2001.1262>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Ma, C.-H., Gao, Z.-J., Zhang, J.-J., Zhang, W., Shao, J.-H., Hai, M.-R., Chen, J.-W., Yang, S.-C., & Zhang, G.-H. (2016). Candidate genes involved in the biosynthesis of triterpenoid saponins in *Platycodon grandiflorum* identified by transcriptome analysis. *Frontiers in Plant Science*, 7, 673. <https://doi.org/10.3389/fpls.2016.00673>
- Ma, Y. X., Fu, H. Z., Li, M., Sun, W., Xu, B., & Cui, J. R. (2007). An anticancer effect of a new saponin component from *Gymnocladus chinensis* Baillon through inactivation of nuclear factor- κ B. *Anti-Cancer Drugs*, 18, 41–46. <https://doi.org/10.1097/O1.cad.0000236317.04199.69>
- Mangas, S., Bonfill, M., Osuna, L., Moyano, E., Tortoriello, J., Cusido, R. M., Teresa Piñol, M., & Palazón, J. (2006). The effect of methyl jasmonate on triterpene and sterol metabolisms of *Centella asiatica*, *Ruscus aculeatus* and *Galphimia glauca* cultured plants. *Phytochemistry*, 67, 2041–2049. <https://doi.org/10.1016/j.phytochem.2006.06.025>
- Menke, F. L., Champion, A., Kijne, J. W., & Memelink, J. (1999). A novel jasmonate-and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate-and elicitor-inducible AP2-domain transcription factor, ORCA2. *The EMBO Journal*, 18, 4455–4463. <https://doi.org/10.1093/emboj/18.16.4455>
- Minoche, A. E., Dohm, J. C., & Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*, 12, 1–15. <https://doi.org/10.1186/gb-2011-12-11-r112>
- Mu, L. H., Gong, Q. Q., Zhao, H. X., Liu, P., & Xi, J. (2010). Triterpenoid Saponins from *Ardisia gigantifolia*. *Chemical and Pharmaceutical Bulletin*, 58, 1248–1251. <https://doi.org/10.1248/cpb.58.1248>
- Nagata, N., Suzuki, M., Yoshida, S., & Muranaka, T. (2002). Mevalonic acid partially restores chloroplast and etioplast development in *Arabidopsis* lacking the non-mevalonate pathway. *Planta*, 216, 345–350. <https://doi.org/10.1007/s00425-002-0871-9>
- Papadopoulou, K., Melton, R. E., Leggett, M., Daniels, M. J., & Osbourn, A. E. (1999). Compromised disease resistance in saponin-deficient plants. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 12923–12928. <https://doi.org/10.1073/pnas.96.22.12923>
- Pérez-Rodríguez, P., Riaño-Pachón, D. M., Corrêa, L. G. G., Rensing, S. A., Kersten, B., & Mueller-Roeber, B. (2010). PlnTFDB: Updated content and new features of the plant transcription factor database. *Nucleic Acids Research*, 38, D822–D827. <https://doi.org/10.1093/nar/gkp805>
- Ponou, B. K., Barboni, L., Teponno, R. B., Mbiatcha, M., Nguefack, T. B., Park, H.-J., Lee, K.-T., & Tapondjou, L. A. (2008). Polyhydroxyoleanane-type triterpenoids from *Combretum molle* and their anti-inflammatory activity. *Phytochemistry Letters*, 1, 183–187. <https://doi.org/10.1016/j.phytol.2008.09.002>
- Rahimi, S., Kim, Y. J., & Yang, D. C. (2015). Production of ginseng saponins: Elicitation strategy and signal transductions. *Applied Microbiology and Biotechnology*, 99, 6987–6996. <https://doi.org/10.1007/s00253-015-6806-8>
- Rai, A., Yamazaki, M., Takahashi, H., Nakamura, M., Kojoma, M., Suzuki, H., & Saito, K. (2016). RNA-seq transcriptome analysis of *Panax japonicus*, and its comparison with other *Panax* species to identify potential genes involved in the saponins biosynthesis. *Frontiers in Plant Science*, 7, 481. <https://doi.org/10.3389/fpls.2016.00481>
- Richter, B. G., & Sexton, D. P. (2009). Managing and analyzing next-generation sequence data. *PLoS Computational Biology*, 5, e1000369. <https://doi.org/10.1371/journal.pcbi.1000369>
- Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., & Creelman, R. (2000). Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*, 290, 2105–2110. <https://doi.org/10.1126/science.290.5499.2105>
- Rodríguez-Concepción, M., Forés, O., Martínez-García, J. F., González, V., Phillips, M. A., Ferrer, A., & Boronat, A. (2004). Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during *Arabidopsis* seedling development. *The Plant Cell*, 16, 144–156. <https://doi.org/10.1105/tpc.016204>
- Rudell, D. R., Mattheis, J. P., Fan, X., & Fellman, J. K. (2002). Methyl jasmonate enhances anthocyanin accumulation and modifies production of phenolics and pigments in Fuji apples. *Journal of the American Society for Horticultural Science*, 127, 435–441. <https://doi.org/10.21273/JASHS.127.3.435>
- Sangwan, R. S., Chaurasiya, N. D., Lal, P., Misra, L., Tuli, R., & Sangwan, N. S. (2008). Withanolide A is inherently de novo biosynthesized in roots of the medicinal plant *Ashwagandha* (*Withania somnifera*). *Physiologia Plantarum*, 133, 278–287. <https://doi.org/10.1111/j.1399-3054.2008.01076.x>
- Sawai, S., & Saito, K. (2011). Triterpenoid biosynthesis and engineering in plants. *Frontiers in Plant Science*, 2, 25. <https://doi.org/10.3389/fpls.2011.00025>
- Sparg, S. G., Light, M. E., & Van Staden, J. (2004). Biological activities and distribution of plant saponins. *Journal of Ethnopharmacology*, 94, 219–243. <https://doi.org/10.1016/j.jep.2004.05.016>
- Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Reviews Genetics*, 20, 631–656. <https://doi.org/10.1038/s41576-019-0150-2>
- Su, S. L., Li, Y. H., OuYang, Z., & Chen, J. (2003). Research progress of triterpene saponins in medicinal plants of the genus *Aralia*. *Journal of Chinese Medicinal Materials*, 26, 144–148.
- Su, W. H., Zhang, G. F., Li, X. H., & Ou, X. K. (2005). Relationship between accumulation of secondary metabolism in medicinal plant and environmental condition. *Chinese Traditional and Herbal Drugs*, 9, 1415–1418.
- Sun, H., Li, F., Xu, Z., Sun, M., Cong, H., Qiao, F., & Zhong, X. (2017). De novo leaf and root transcriptome analysis to identify putative genes involved in triterpenoid saponins biosynthesis in *Hedera helix* L. *PLoS One*, 12, e0182243. <https://doi.org/10.1371/journal.pone.0182243>
- Sun, H. X., Xie, Y., & Ye, Y. P. (2009). Advances in saponin-based adjuvants. *Vaccine*, 27, 1787–1796. <https://doi.org/10.1016/j.vaccine.2009.01.091>
- Sun, H. X., Ye, Y. P., Pan, H. J., & Pan, Y. J. (2004). Adjuvant effect of *Panax notoginseng* saponins on the immune responses to ovalbumin in mice. *Vaccine*, 22, 3882–3889. <https://doi.org/10.1016/j.vaccine.2004.04.012>
- Sun, S. F., Jiang, X., Zhang, C. L., & Zhang, Y. Y. (2017). Progress in research on chemical component and pharmacological activities of *Ardisia* plants. *Shandong Chemical Industry Journal*, 46, 77–79.
- Tang, Y. (2007). Clinical observation on treatment of rheumatoid arthritis with Chinese herbs. *Journal of Sichuan Traditional Chinese Medicine*, 25, 54–55.
- Tatusov, R. L. (2001). The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*, 29, 22–28. <https://doi.org/10.1093/nar/29.1.22>

- Van Der Fits, L., & Memelink, J. (2001). The jasmonate-inducible AP2/ERF-domain transcription factor ORCA3 activates gene expression via interaction with a jasmonate-responsive promoter element. *The Plant Journal*, *25*, 43–53.
- Vermeersch, M., Foubert, K., Inocência Da Luz, R., Puyvelde, L. V., Pieters, L., Cos, P., & Maes, L. (2009). Selective antileishmania activity of 13,28-Epoxy-oleanane and related triterpene saponins from the plant families Myrsinaceae, Primulaceae, Aceraceae and Icacinaeae. *Phytotherapy Research*, *23*, 1404–1410. <https://doi.org/10.1002/ptr>
- Vranová, E., Coman, D., & Grussem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology*, *64*, 665–700. <https://doi.org/10.1146/annurev-arplant-050312-120116>
- Wang, L., Feng, Z., Wang, X. I., Wang, X., & Zhang, X. (2010). Degseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, *26*(1), 136–138. <https://doi.org/10.1093/bioinformatics/btp612>
- Wang, S., Wang, X., He, Q., Liu, X., Xu, W., Li, L., Gao, J., & Wang, F. (2012). Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. *Plant Cell Reports*, *31*, 1437–1447. <https://doi.org/10.1007/s00299-012-1259-3>
- Wang, X. F., Gao, W. Q., Liu, J. F., Ni, Y. Y., & Jiang, Z. P. (2015). Plant defensive strategies and environment-driven mechanisms. *Chinese Journal of Ecology*, *34*, 3542–3552.
- Wang, X. W., Luan, J. B., Li, J. M., Bao, Y. Y., Zhang, C. X., & Liu, S. S. (2010). De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, *11*, 400. <https://doi.org/10.1186/1471-2164-11-400>
- Wei, R. (2017). *The influence of environmental factors on contents of biomass and saponins of Ardisia gigantifolia*. [master's thesis]. [Guangzhou]: Zhongkai University of Agriculture and Engineering.
- Wei, R. (2018). Community characteristics and conservation strategy of *Ardisia gigantifolia* in Nanxiong, Guangdong Province. *Chinese Traditional and Herbal Drugs*, 1430–1436.
- Wei, R., Yun, J. Q., Xie, S. M., & Chen, H. F. (2015). Research progress on the resources and applications of *Ardisia gigantifolia*. *Guangdong Forestry Science and Technology*, *31*, 94–98.
- Werner, T., & Schmülling, T. (2009). Cytokinin action in plant development. *Current Opinion in Plant Biology*, *12*, 527–538. <https://doi.org/10.1016/j.pbi.2009.07.002>
- Xiang, C. L., & Feng, Z. J. (2002). Medicinal plant resources of Myrsinaceae in Guangdong Province. *Subtropical Plant Science*, *31*, 52–55.
- Yan, M. C., Liu, Y., Chen, H., Ke, Y., Xu, Q. C., & Cheng, M. S. (2006). Synthesis and antitumor activity of two natural N-acetylglucosamine-bearing triterpenoid saponins: Lotoidoside D and E. *Bioorganic & Medicinal Chemistry Letters*, *16*, 4200–4204. <https://doi.org/10.1016/j.bmcl.2006.05.086>
- Yang, C. Q., Fang, X., Wu, X. M., Mao, Y. B., Wang, L. J., & Chen, X. Y. (2012). Transcriptional regulation of plant secondary metabolism. *Journal of Integrative Plant Biology*, *54*, 703–712.
- Yang, J. (2015). *Studies on the exploitation and utilization of medicinal plant Ardisia crenata Sims* [master's thesis]. [Sichuan]: Sichuan Agricultural University.
- Ye, J., Mao, D., Cheng, S., Zhang, X., Tan, J., Zheng, J., & Xu, F. (2020). Comparative transcriptome analysis reveals the potential stimulatory mechanism of terpene trilactone biosynthesis by exogenous salicylic acid in *Ginkgo biloba*. *Industrial Crops and Products*, *145*, 112104. <https://doi.org/10.1016/j.indcrop.2020.112104>
- Yendo, A. C. A., de Costa, F., Da Costa, C. T., Colling, L. C., Gosmann, G., & Fett-Neto, A. G. (2014). Biosynthesis of plant triterpenoid saponins: Genes, enzymes and their regulation. *Mini-Reviews in Organic Chemistry*, *11*, 292–306.
- Yu, Z. X., Li, J. X., Yang, C. Q., Hu, W. L., Wang, L. J., & Chen, X. Y. (2012). The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Molecular Plant*, *5*, 353–365. <https://doi.org/10.1093/mp/ssr087>
- Zhang, Q. H. (1994). A survey of the phytochemical constituents of The genus *Ardisia*. *West China Journal of Pharmaceutical Sciences*, *9*, 99–103.
- Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., Banf, M., Dai, X., Martin, G. B., Giovannoni, J. J., & Zhao, P. X. (2016). iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Molecular Plant*, *9*, 1667–1670.
- Zhou, Z. J. (2017). Progress and prospect of research on chemical components and pharmacological activity of *Ardisia gigantifolia* Stapf. *Journal of Guangxi Agriculture*, *32*, 50–53.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Lei, Y., Harris, A. J., Wang, A., Zhao, L., Luo, M., Li, J., & Chen, H. (2022). Comparative transcriptomic analysis of genes in the triterpene saponin biosynthesis pathway in leaves and roots of *Ardisia kteniophylla* A. DC., a plant used in traditional Chinese medicine. *Ecology and Evolution*, *12*, e8920. <https://doi.org/10.1002/ece3.8920>